

Development of DArT markers for a linkage map of flue-cured tobacco

LU XiuPing^{1,2}, GUI YiJie^{1*}, XIAO BingGuang^{2*}, LI YongPing², TONG ZhiJun¹, LIU Yun¹, BAI XueFei¹, WU WeiRen¹, XIA Ling³, HUTTNER Eric³, KILIAN Adrzej³ & FAN LongJiang¹

¹ Department of Agronomy, Zhejiang University, Hangzhou 310058, China;

² Yunnan Academy of Tobacco Agricultural Sciences and China Tobacco Breeding Research Center at Yunnan, Yuxi 653100, China;

³ DArT P/L, PO Box 7141, Yarralumla, Canberra, ACT 2600, Australia

Received April 25, 2012; accepted July 20, 2012; published online October 29, 2012

Tobacco (*Nicotiana tabacum*) is one of the most economically important nonfood crops, and flue-cured tobacco accounts for approximately 80% of world tobacco production. An extremely narrow genetic diversity in the tobacco pool has led to a low efficiency of PCR-based molecular markers (such as AFLP and SSR). Diversity Arrays Technology (DArT) is a high-throughput hybridisation-based marker system that has been developed in many plants including wheat, which, like tobacco, has a complex genome. In this study, we developed a tobacco DArT chip that included 7680 representative sequence tags based on typical tobacco accessions. The 1076 DArT markers of flue-cured tobacco were identified and most (82.1%) of their polymorphism information contents (PICs) were greater than 0.4. An integrated linkage map that included 851 markers (238 DArT and 613 SSR), which is the highest density map of flue-cured tobacco to date, was constructed. This chip-based DArT system provides an alternative in high-throughput marker genotyping for tobacco.

***Nicotiana tabacum*, diversity arrays technology (DArT), linkage map, DH population, molecular marker**

Citation: Lu X P, Gui Y J, Xiao B G, et al. Development of DArT markers for a linkage map of flue-cured tobacco. *Chin Sci Bull*, 2013, 58: 641–648, doi: 10.1007/s11434-012-5453-z

Tobacco (*Nicotiana tabacum*) is one of the most economically important nonfood crops in the world, and several types (such as flue-cured, burley, oriental and cigar) have been domesticated/improved. As the most important type, flue-cured tobacco accounts for approximately 80% of world tobacco production (Universal Leaf Tobacco Company, 2012). Tobacco was domesticated from only a few wild progenitors and their genetic diversity has been decreasing in recent decades under strict breeding selection [1,2], which results in an extremely narrow genetic diversity in the flue-cured tobacco pool as revealed by several molecular marker-based studies on diversity [1,3,4].

Molecular markers play an important role in genetic diversity, linkage mapping, QTL mapping and marker-assisted

crop improvement. From the first molecular marker restriction fragment length polymorphism (RFLP) reported by Botstein et al. [5] in the 20th century, diverse molecular markers, such as randomly amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism markers (AFLPs), have been developed and used in crops. As in other crops, RFLP, RAPD and AFLP markers have been developed and used in tobacco since the 1990's [6–13]. Recently, simple sequence repeat (SSR) [1–3,14,15] and other new types of markers, such as small RNA-based molecular markers [16], have been further developed and used in studies of diversity and linkage mapping in tobacco. Based on the above molecular marker systems, efforts to create a linkage map in tobacco began in this century. To date, the most important progress in this field has been an SSR-based linkage map based on an F₂ population derived

*Corresponding authors (email: yijiegui@zju.edu.cn; xiaobg@263.net)

from two different types of parental tobacco lines (Hicks Broad Leaf and Red Russia) by Bindler et al. [14,15]. The first linkage map of flue-cured tobacco based on a double haploid (DH) population from a cross between Speight G-28 and NC2326 was constructed using AFLP and inter SSR markers in 2006 [17], and no further maps have been available for flue-cured tobacco. As mentioned above, a very narrow genetic diversity has been observed in flue-cured tobacco, which makes the development of PCR-based molecular markers, in general, very inefficient, and a more sensitive (usually also high cost) marker scanning protocol (such as that described by Bindler et al. [15]) or a high-throughput marker system is required.

Diversity Arrays Technology (DArT) is developed as a high-throughput hybridisation-based marker system which captures the values of the parallel nature of a microarray platform without relying on DNA sequence information [18]. DArT markers have been developed and used in many plants, including wheat and barley, which, like tobacco, have complex genomes, and their genome sequences are not yet available for genetic mapping and studies of genetic diversity [19–22]. This technology combines a reduction in genomic complexity with hybridization-based polymorphism detection using high throughput, solid-state platforms and has the potential to generate hundreds of high-quality genomic dominant markers with a cost-effective and time-competitive trade-off.

In this study, we developed a tobacco DArT chip that includes 7680 representative sequence tags based on four main types of tobacco accessions and further constructed a flue-cured tobacco linkage map with the highest reported density of molecular markers using the DArT arrays.

1 Materials and methods

1.1 Plant materials and DNA extraction

A panel of 10 typical tobacco varieties were selected and used in this study (Table 1). Five typical tobacco varieties from 4 types (flue-cured: HHDJY and Hicks Broad Leaf; varieties of burley, oriental and cigar) were used to prepare DArT libraries. An F_1 -derived DH population obtained from a cross between two flue-cured parental lines, HHDJY and Hicks Broad Leaf, with 207 progenic lines was used for construction of the linkage map. Total genomic DNA was extracted from fresh leaf tissue (~200 mg) of the two parent lines and DH population lines using the CTAB protocol [23].

1.2 Development of DArT for tobacco

(i) Preparation of genomic representations. Preliminary tests of various methods for complexity reduction and library representations were performed by two approaches (*PstI/TaqI* and *PstI/HpaII*) for the development of DArT

clone libraries for tobacco. Details of this methodology have been described previously [18,19] and are presented briefly below. Five cultivars representing main types of tobacco were selected for library development (Table 1). The representations were generated by cutting 100 ng DNA samples with *PstI* and a frequent cutter (*TaqI* or *HpaII*) (NEB). A *PstI* adapter (5'-CACGATGGATCCAGTGCA-3' annealed with 5'-CTGGATCCATCGTGCA-3') was simultaneously ligated with T4 DNA ligase (NEB). Two libraries of clones were developed for each discovery array. The first array that contained 3072 clones was developed using the *PstI/TaqI* method and the other array was developed using *PstI/HpaII* and contained 4608 clones.

(ii) Microarray preparation. A library was prepared from the amplification products as described previously [24]. The amplification reactions were dried at 37°C, washed with 70% ethanol, dissolved in DArT Spotter 2 spotting buffer, and then printed on polylysine-coated slides (Erie Scientific, Portsmouth, NH, USA) using a MicroGrid II arrayer (Biorobotics, Cambridge, UK). After printing, slides were placed in a water bath for 2 min at 95°C for DNA denaturation, and then dried by centrifugation at 500×g for 7 min.

(iii) Genotyping of DNA samples. Genotyping was performed essentially as described previously [18]. Genomic representations were generated from the 10 tobacco varieties and DH lines using the same complexity reduction method that was used for library construction (*PstI/TaqI*, *PstI/HpaII* or *PstI/MspI*). *HpaII* and *MspI* are isoschizomers that have different sensitivities to the methylation status of CCGG tetra nucleotides. The resulting genomic representation was labelled with fluorescent nucleotides (targets) and hybridised on a microarray printed with the DArT clones overnight at 65°C.

(iv) Slide scanning, image analysis and polymorphism scoring. After hybridization, the slides were washed, dried and scanned as described previously [18]. Three images were generated from each slide. One image was produced with a 488-nm laser (FAM labelled). The remaining two images were used as independent targets; one was produced with a 543-nm laser (Cy3-labelled targets) and the other was produced with a 633-nm laser (Cy5-labelled targets). Image-processing and the polymorphism calculation were performed using DArTsoft version 7.4.6, a dedicated software package developed at DArT P/L. The improved software localized spots, rejected those with weak reference signals, normalized relative hybridization intensities ($=\log[\text{Cy3 or Cy5 targets}]/[\text{FAM-reference}]$), calculated the median value for replicate spots, identified polymorphic clones by fuzzy C-means (FCM) [25], and classified polymorphic clones as being present (1) or absent (0) in the representation hybridized to a slide. Standard methods of marker discovery were deployed using several quality parameters that were automatically extracted from the array data: (1) *P* value, the variance of the relative target hybridisation intensity between allelic states as a percentage of the total variance;

(2) Call-Rate, the percentage of DNA samples with binary (“0” or “1”) allele calls (percentage of effective scores); and (3) Reproducibility, the fraction of concordant calls for replicate assays. The markers reported in this paper all showed $P > 70$, Call-Rate > 80 , and Reproducibility > 90 .

1.3 Diversity analysis

The genetic relationship of the 7 flue-cured tobacco varieties in Table 1 was investigated using the NTSYSpc 2.10e program [26] with the following settings: Qualitative data, Dice coefficient, Sequential Agglomerative Hierarchical and Nested clustering method (SAHN), Unweighted Pair-Group Method, Arithmetic average (UPGMA).

1.4 Linkage map construction

A genetic map was constructed using JoinMap 4.0 [27]. Based on SSR markers developed in our DH population, SSRs from the 24 tobacco linkage groups reported in [15] were initially used to construct a framework map for our DH population. The remaining markers were then assigned to its groups by searching for linkages with the highest possible LOD score to minimise the number of apparent cross-overs. Grouping of markers was verified using the JoinMap 4.0 program with the following settings: linkages with an REC smaller than 0.400 and LOD larger than 1.00 were used for the removal of loci with respect to jumps in goodness-of-fit 5.000, Kosambi mapping function.

2 Results

2.1 Evaluation of complexity reduction methods and array development

A critical step in DArT development is the preparation of genome representations through the reduction of genome complexity [24]. Based on our previous experiments with *PstI*-based representations in several plant genomes, we initially evaluated several combinations of *PstI* with a frequent restriction enzyme (RE) as an approach to complexity reduction in tobacco. Consequently, DNA samples from five typical tobacco varieties (Table 1) were digested with *PstI* and several common cutting REs (*PstI/AluI*, *PstI/BanII*, *PstI/BstNI*, *PstI/HaeIII*, *PstI/HpaII*, *PstI/MseI*, and *PstI/TaqI*), ligated to a *PstI* adaptor and then amplified with the *PstI*-0 primer. Gel analysis of their PCR products showed that a uniform smear (without major bands) appeared in the *PstI/TaqI* and *PstI/HpaII* combinations, while other RE combinations gave a smear with one or more dominant bands (data not shown). Dominant bands represent highly amplified restriction fragments and correspond to abundant repetitive sequences in the representation. Segments from both the *PstI/TaqI* and *PstI/HpaII* methods were chosen for further library construction.

After a test of DArT performance in a subset of our DH population (which included 20 DH segregation lines), 3072 random clones from the *PstI/TaqI* library and 4608 clones from the *PstI/HpaII* library were selected and printed onto polylysine-coated slides. Ultimately, a tobacco DArT array that included 7680 representative sequence tags was prepared and used for tobacco genotyping.

2.2 Identification of DArT markers for flue-cured tobacco

To identify DArT markers for flue-cured tobacco, in addition to two flue-cured parental lines (HHDJY and Hicks Broad Leaf), 5 other typical flue-cured varieties (Table 1) were scanned using the newly developed DArT chip. A total of 1076 markers were polymorphic among the 7 tobacco varieties based on the DArTsoft analysis. The distribution of polymorphism information content (PIC) values of the 1076 DArT markers indicated that most of them (82.1%) had PIC values between 0.4 and 0.5, and 91.1% of their PIC values were larger than 0.2 (Table 2), similar to the results in other plants; for example, in sorghum [28], 69.5% and 94.7% of DArT markers have PIC values above 0.4 and 0.2, respectively. We also created a phylogenetic tree for the 7 flue-cured tobacco varieties using the 1076 markers (Figure 1). This tree provides a phylogenetic relation similar to those obtained in other studies, such as with an SSR-based investigation [1]. Taken together, the results suggested that the present DArT arrays work well for tobacco.

Table 1 Tobacco varieties used in this study

Variety	Type
HHDJY	Flue-cured
Hicks Broad Leaf	Flue-cured
K326	Flue-cured
Coker85	Flue-cured
Coker176	Flue-cured
C151	Flue-cured
T.I.245	Flue-cured
Florida 301	Cigar
Burley 21	Burley
Turkey Basma	Oriental

Table 2 Polymorphism information content (PIC) values for 1076 DArT markers

PIC value	DArTs	Percentage (%)
0.5–0.4	883	82.1
0.4–0.3	56	5.2
0.3–0.2	41	3.8
0.2–0.1	30	2.8
0.1–0	66	6.1

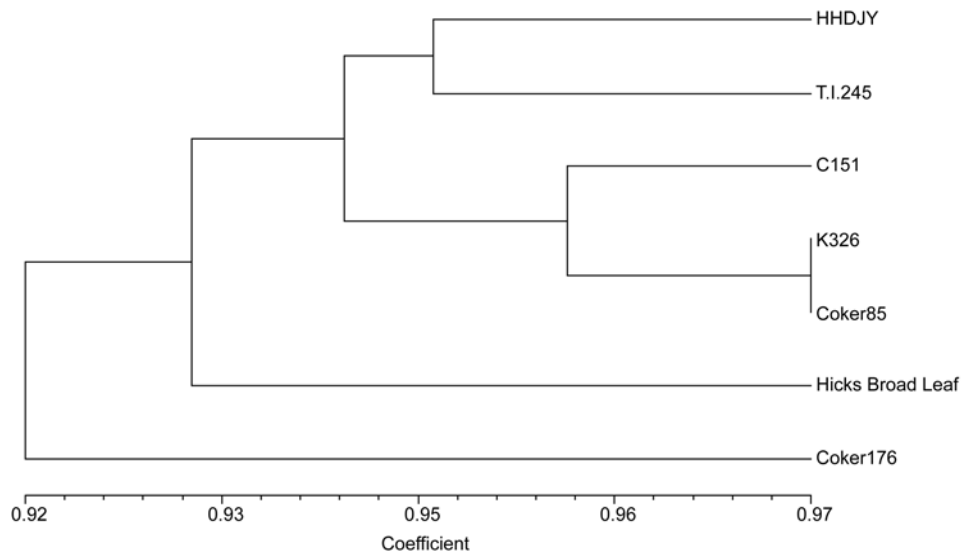


Figure 1 Phylogenetic tree of seven flue-cured tobacco varieties based on 1076 DArT markers.

2.3 An integrated DArT+SSR linkage map of flue-cured tobacco

In the above DArT slide scanning, a total of 317 polymorphic DArT tags were found in two parental flue-cured lines of our DH population. Based on their PIC values (>0.325), 246 DArTs were selected from these 317 markers and used for the construction of a linkage map for the flue-cured tobacco population. At the same time, over 600 SSR markers have been developed for our DH population based on the SSRs described by Bindler et al. [15] and sequences from the Tobacco Genome Initiative (TGI) (<http://www.pngg.org/tgi/>) and RNA-Seq by ourselves, and these were also used to create a DArT+SSR integrated linkage map of flue-cured tobacco.

Finally, an integrated linkage map that included 851 markers (238 DArT and 613 SSR) in 24 linkage groups was obtained (Figure 2). The total length of the genetic map was 2291 cM, with individual linkage groups ranging from 27 to 165 cM, or 15 to 103 markers (Table 3). In the current map, 20 linkage groups have both DArT and SSR markers in each group while the other four linkage groups consisted of only DArT or SSR markers. Of the 613 SSR markers in our map, 218 were developed from the SSR markers described by Bindler et al. [15], which are scattered into 23 linkage groups. Alignment of our map with theirs showed a high level of conservation of the order of the 176 SSR loci, with a few exceptions, which might be due to small chromosome structural inversions between the parental lines used by us and Bindler et al. [15].

3 Discussion

In this study, we developed a DArT marker system for to-

bacco and demonstrated that this system can provide high-quality markers for diversity analyses and genetic linkage mapping of flue-cured tobacco, the most important type of tobacco in China and many other countries. Chip-based marker systems, such as SNP chip and DArT, offer an easy way to scan particular loci/tags in target lines, and their results usually can present a basic genetic background or characteristics for those lines and therefore can provide solid data regarding their genetic diversity, linkage mapping, molecular marker-assisted breeding, etc. DArT clones can be sequenced and converted into sequence tags or PCR-based markers, which should be very helpful for linkage map-based genomic assembly and molecular breeding. In short, even in the era of high-throughput sequencing, chip-based systems like DArT still play important roles in high-throughput marker analysis or genotyping.

In our integrated linkage map, one linkage group was constructed with only DArT markers (Group 5) and no DArT markers were used in other three groups (groups 1, 4, and 16) (Table 3). Since there are 238 DArT markers in the linkage map, or an average of 10 markers in each linkage group, the distribution of DArT loci in the tobacco genome does not appear to be uniform. A similar situation has been observed in other plants [29]. This bias may be the result of several factors, such as repeats and methylation. Restriction endonucleases play a key role in the collection of representative segments for DArT tags. In this study, we used the uncommon cutter *PstI* in DArT development and the genomic distribution of the cutter's sites should determine the distribution of DArT tags. Tobacco has a large and repetitive genome [30]. The distribution of repetitive sequences will dramatically change the output of DArT tags. Similarly, *PstI* is sensitive to methylation and the methylation level of the cutter's sites in the tobacco genome should also influence the output of DArT tags.

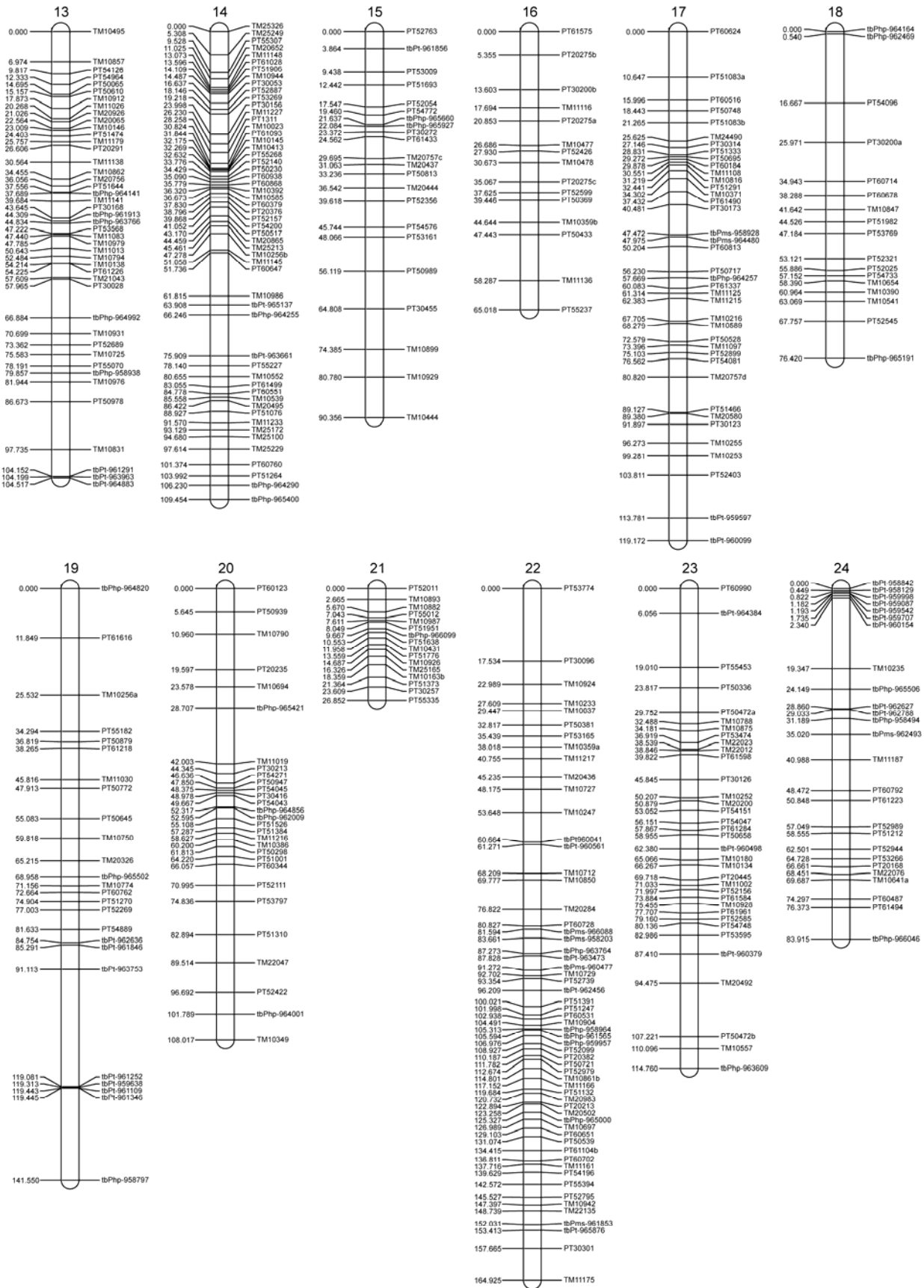


Figure 2 Genetic linkage map of flue-cured tobacco based on a DH population (HHDJY×Hicks Broad Leaf). Map distance (cM) and marker name are shown on the left and right sides of each linkage group, respectively. SSR marker names begin with the prefix “PT” or “TM” while DaRT names start with the prefix “tb”.

Table 3 Summary of the DArT+SSR linkage map of flue-cured tobacco

Linkage group	SSR	DArT	Total markers	Total length (cM)	Average marker distance	Maximal marker distance
1	22	0	22	92	4.2	10.8
2	58	2	60	120	2.0	9.9
3	37	9	46	68	1.5	7.1
4	33	0	33	62	1.9	8.1
5	0	103	103	148	1.4	26.2
6	19	3	22	89	4.1	8.3
7	27	6	33	104	3.2	27.4
8	23	3	26	55	2.1	6.0
9	29	3	32	120	3.8	11.9
10	17	3	20	76	3.8	8.5
11	19	5	24	80	3.3	16.0
12	16	29	45	91	2.0	13.2
13	36	8	44	105	2.4	11.0
14	49	5	54	109	2.0	10.1
15	19	3	22	90	4.1	9.6
16	15	0	15	65	4.3	10.8
17	34	5	39	119	3.1	10.6
18	14	3	17	76	4.5	16.1
19	15	10	25	142	5.7	28.0
20	25	4	29	108	3.7	13.3
21	15	1	16	27	1.7	3.2
22	45	14	59	165	2.8	17.5
23	31	4	35	115	3.3	13.0
24	13	13	26	84	3.2	17.0

This study led to a linkage map with the highest density of molecular markers to date for flue-cured tobacco. As mentioned above, an extremely narrow genetic diversity due to a domestication bottleneck and recent strict improvement was observed in flue-cured tobacco, which impedes the development of molecular markers and the construction of linkage maps in flue-cured tobacco. Tobacco is tetraploid, which further complicates efforts to produce a linkage map in tobacco. This could explain why only 238 DArT markers (3.1%) of the 7680 tags in the DArT chip were developed and grouped into the linkage map.

This work was supported by Yunnan Provincial Tobacco Company (08A05/2011YN04) and China National Tobacco Company (110200701023/110201201003(JY-03)).

- 1 Moon H S, Nicholson J S, Heineman A, et al. Changes in genetic diversity of U.S. flue-cured tobacco germplasm over seven decades of cultivar development. *Crop Sci*, 2009, 49: 498–508
- 2 Moon H S, Nifong J M, Nicholson J S, et al. Microsatellite-based analysis of tobacco (*Nicotiana tabacum* L.) genetic resources. *Crop Sci*, 2009, 49: 2149–2159
- 3 Zhang H Y, Liu X Z, Li T S, et al. Genetic diversity among flue-

- 4 cured tobacco (*Nicotiana tabacum* L.) revealed by amplified fragment length polymorphism. *Bot Stud*, 2006, 47: 223–229
- 5 Yang B C, Xiao B G, Chen X J, et al. Assessing the genetic diversity of tobacco germplasm using intersimple sequence repeat and inter-retrotransposon amplification polymorphism markers. *Annu Appl Biol*, 2007, 150: 393–401
- 6 Botstein D, White R L, Skolnick M, et al. Construction of a genetic-linkage map in man using restriction fragment length polymorphisms. *Am J Human Genet*, 1980, 32: 314–331
- 7 Coussirat J C. Genetic diversity and varietal identification in *Nicotiana tabacum* with RAPD markers. *Annual du Tabac Section 2*, 1994
- 8 Bai D, Reeleder R, Brandle J E. Identification of two RAPD markers tightly linked with the *Nicotiana debneyi* gene for resistance to black root rot of tobacco. *Theor Appl Genet*, 1995, 91: 1184–1189
- 9 Lin T Y, Kao Y Y, Lin S, et al. A genetic linkage map of *Nicotiana plumbaginifolia*/*Nicotiana longiflora* based on RFLP and RAPD markers. *Theor Appl Genet*, 2001, 103: 905–911
- 10 Ren N, Timko M P. AFLP analysis of genetic polymorphism and evolutionary relationships among cultivated and wild nicotiana species. *Genome*, 2001, 44: 559–571
- 11 Johnson E S, Wolff M F, Wernsman E A, et al. Marker-assisted selection for resistance to black shank disease in tobacco. *Plant Dis*, 2002, 86: 1303–1309
- 12 Lewis R S, Milla S R, Levin J S. Molecular and genetic characterization of *Nicotiana tabacum* L. Chromosome segments in tobacco mosaic virus-resistant tobacco accessions. *Crop Sci*, 2005, 45: 2355–2362
- 13 Julio E, Verrier J L, de Borne F D. Development of SCAR markers

- linked to three disease resistances based on AFLP within *Nicotiana tabacum* L. Theor Appl Genet, 2006, 112: 335–346
- 13 Moon H, Nicholson J S. AFLP and SCAR markers linked to tomato spotted wilt virus resistance in tobacco. Crop Sci, 2007, 47: 1887–1894
 - 14 Bindler G, van der Hoeven R, Gunduz I, et al. A microsatellite marker based linkage map of tobacco. Theor Appl Genet, 2007, 114: 341–349
 - 15 Bindler G, Plieske J, Bakaher N, et al. A high density genetic map of tobacco (*Nicotiana tabacum* L.) obtained from large-scale microsatellite marker development. Theor Appl Genet, 2011, 123: 219–230
 - 16 Gui Y J, Yan G H, Bo S P, et al. iSNAP: A small RNA-based molecular marker technique. Plant Breed, 2011, 130: 515–520
 - 17 Xiao B G, Xu Z L, Chen X J, et al. Genetic linkage map constructed by using a DH population for flue-cured tobacco. Acta Tabacaria Sin, 2006, 35–40
 - 18 Jaccoud D, Peng K M, Feinstein D, et al. Diversity arrays: A solid state technology for sequence information independent genotyping. Nucleic Acids Res, 2001, 29: 7
 - 19 Wenzl P, Carling J, Kudrna D, et al. Diversity arrays technology (DArT) for whole-genome profiling of barley. Proc Natl Acad Sci USA, 2004, 101: 9915–9920
 - 20 Akbari M, Wenzl P, Caig V, et al. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet, 2006, 113: 1409–1420
 - 21 Tinker N A, Kilian A, Wight C P, et al. New DArT markers for oat provide enhanced map coverage and global germplasm characterization. BMC Genomics, 2009, 10: 39
 - 22 Alsop B P, Farre A, Wenzl P, et al. Development of wild barley-derived DArT markers and their integration into a barley consensus map. Mol Breed, 2011, 27: 77–92
 - 23 Sambrook J, Russell D W. Molecular Cloning: A Laboratory Manual (Chinese version translated by Huang P T). 3rd Ed. Beijing: Science Press, 2001. 547–610
 - 24 Xia L, Peng K M, Yang S Y, et al. DArT for high-throughput genotyping of cassava (*Manihot esculenta*) and its wild relatives. Theor Appl Genet, 2005, 110: 1092–1098
 - 25 Bezdek J C, Ehrlich R, Full W. FCM—The fuzzy C-means clustering- algorithm. Comp Geosci, 1984, 10: 191–203
 - 26 Rohlf F J. NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System. New York: Exeter Software, 2000
 - 27 Van Ooijen J W. Joinmap[®] 4, Software for The Calculation of Genetic Linkage Maps in Experimental Population. Wageningen: Kyazma B V, 2006
 - 28 Mace E S, Xia L, Jordan D R, et al. DArT markers: Diversity analyses and mapping in sorghum bicolor. BMC Genomics, 2008, 9: 26
 - 29 Hearnden P R, Eckermann P J, McMichael G L, et al. A genetic map of 1000 SSR and DArT markers in a wide barley cross. Theor Appl Genet, 2007, 115: 383–391
 - 30 Opperman C H, Lommel S A. The Tobacco Genome Initiative: Gene discovery and data mining in *Nicotiana tabacum*. Plant & Animal Genomes XI Conference, San Diego, CA, 2007

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.